AMENDMENTS TO THE CLAIMS:

Claims 1-40 (Cancelled)

41. (Currently amended) A method of classifying new datasets within a predetermined number of categories based on assignment of a plurality of sample datasets to each category, the method comprising the machine-executed steps:

constructing a trainable semantic vector for each sample dataset relative to the predetermined categories in a multi-dimensional semantic space;

constructing a trainable semantic vector for each category based on the trainable semantic vectors for the sample datasets;

receiving a new dataset;

constructing a trainable semantic vector for the new dataset;

determining a distance between the trainable semantic vector for the new dataset and the trainable semantic vector of each category; and

classifying the new dataset within the category whose trainable semantic vector has the shortest distance to the trainable semantic vector of the new ~~dataset.~~dataset;

wherein:

the new data set or each of the sample data sets includes at least one data point;

each data point corresponds to at least one of a word, a phrase, a sentence, a color, a typography, a punctuation, a picture, and a character string; and

the trainable semantic vector for each sample data set or the new dataset is constructed by performing the steps of:

for each data point, identifying a relationship between each data point and predetermined categories corresponding to dimensions in the semantic space;

determining the significance of each data point with respect to the predetermined categories;

constructing a trainable semantic vector for each data point, wherein each trainable semantic vector has dimensions equal to the number of predetermined categories and represents the relative strength of its corresponding data point with respect to each of the predetermined categories; and

combining the trainable semantic vector for each of the at least one data point to form the semantic vector of the sample dataset or the new dataset.

42. (Original) The method of Claim 41 wherein the datasets correspond to documents.

43. (Original) The method of Claim 41 wherein the datasets correspond to email messages and the categories correspond to frequently asked questions with substantially static responses.

44. (Original) The method of Claim 41, further comprising the steps:

detecting when a prescribed number of new datasets has been classified; and

updating the trainable semantic vectors for each of the categories.

45. (Original) The method of Claim 44, wherein the step of updating comprises the step of re-constructing trainable semantic vectors for each category based on the trainable semantic vectors for the sample datasets and the trainable semantic vectors for the new datasets added to each category.

46. (Currently amended l) A method of classifying new datasets within a predetermined number of categories based on assignment of a plurality of sample datasets to each category, the method comprising the machine-executed steps:

constructing a trainable semantic vector for each sample dataset relative to the predetermined categories in a multi-dimensional semantic space;

receiving a new dataset;

constructing a trainable semantic vector for the new dataset;

identifying a select number of sample datasets whose trainable semantic vectors are closest in distance to the trainable semantic vector for the new dataset; and

classifying the new dataset in the category containing the greatest number of the select sample ~~datasets.~~datasets;

wherein:

the new data set or each of the sample data sets includes at least one data point;

each data point corresponds to at least one of a word, a phrase, a sentence, a color, a typography, a punctuation, a picture, and a character string; and

the trainable semantic vector for each sample data set or the new dataset is constructed by performing the steps of:

for each data point, identifying a relationship between each data point and predetermined categories corresponding to dimensions in the semantic space;

determining the significance of each data point with respect to the predetermined categories;

constructing a trainable semantic vector for each data point, wherein each trainable semantic vector has dimensions equal to the number of predetermined categories and represents the relative strength of its corresponding data point with respect to each of the predetermined categories; and

combining the trainable semantic vector for each of the at least one data point to form the semantic vector of the sample dataset or the new dataset.

47. (Original) The method of Claim 46 wherein the datasets correspond to documents.

48. (Original) The method of Claim 46 wherein the datasets correspond to email messages and the categories correspond to frequently asked questions with substantially static responses.49.

49. (Original) The method of Claim 46, further comprising the steps:

detecting when a prescribed number of new datasets has been classified; and

adding the new datasets to the set of sample datasets.

50. (Currently amended) A method of classifying new datasets within a predetermined number of categories, the method comprising the machine-executed steps:

receiving a new dataset;

constructing a trainable semantic vector for the new dataset, where the dimensions of the trainable semantic vector correspond to the predetermined number of categories;

classifying the dataset in the category whose corresponding dimension in the trainable semantic vector has the largest ~~value.~~value;

wherein:

the new data set includes one or more data point;

each data point corresponds to at least one of a word, a phrase, a sentence, a color, a typography, a punctuation, a picture, and a character string; and

the trainable semantic vector for the new dataset is constructed by performing the steps of:

for each data point within the new dataset, identifying a relationship between each data point and predetermined categories corresponding to dimensions in the semantic space;

determining the significance of each data point with respect to the predetermined categories;

constructing a trainable semantic vector for each data point, wherein each trainable semantic vector has dimensions equal to the number of predetermined categories and represents the relative strength of its corresponding data point with respect to each of the predetermined categories; and

combining the trainable semantic vector for each data point to form the semantic vector of the new dataset.

51. (Original) The method of Claim 50 wherein the datasets correspond to documents.

52. (Original) The method of Claim 50 wherein the datasets correspond to email messages and the categories correspond to frequently asked questions with substantially static responses.

Claims 53-62 (Cancelled)

63. (Currently amended) A system for classifying new datasets within a predetermined number of categories based on assignment of a plurality of sample datasets to each category, the system comprising:

a computer configure to:

construct a trainable semantic vector for each sample dataset relative to the predetermined categories in a multi-dimensional semantic space;

construct a trainable semantic vector for each category based on the trainable semantic vectors for the sample datasets;

receive a new dataset;

construct a trainable semantic vector for the new dataset;

determine a distance between the trainable semantic vector for the new dataset and the trainable semantic vector of each category; and

classify the new dataset within the category whose trainable semantic vector has the shortest distance to the trainable semantic vector of the new dataset;~~dataset.~~

wherein:

the new data set or each of the sample data sets includes at least one data point;

each data point corresponds to at least one of a word, a phrase, a sentence, a color, a typography, a punctuation, a picture, and a character string; and

the trainable semantic vector for each sample data set or the new dataset is constructed by performing the steps of:

for each data point, identifying a relationship between each data point and predetermined categories corresponding to dimensions in the semantic space;

determining the significance of each data point with respect to the predetermined categories;

constructing a trainable semantic vector for each data point, wherein each trainable semantic vector has dimensions equal to the number of predetermined categories and represents the relative strength of its corresponding data point with respect to each of the predetermined categories; and

combining the trainable semantic vector for each of the at least one data point to form the semantic vector of the sample dataset or the new dataset.

64. (Currently amended) A system for classifying new datasets within a predetermined number of categories based on assignment of a plurality of sample datasets to each category, the system comprising:

a computer configured to:

construct a trainable semantic vector for each sample dataset relative to the predetermined categories in a multi-dimensional semantic space;

receive a new dataset;

construct a trainable semantic vector for the new dataset;

identify a select number of sample datasets whose trainable semantic vectors are closest in distance to the trainable semantic vector for the new dataset; and

classify the new dataset in the category containing the greatest number of the select sample ~~datasets~~datasets;

wherein:

the new data set or each of the sample data sets includes at least one data point;

each data point corresponds to at least one of a word, a phrase, a sentence, a color, a typography, a punctuation, a picture, and a character string; and

the trainable semantic vector for each sample data set or the new dataset is constructed by performing the steps of:

for each data point, identifying a relationship between each data point and predetermined categories corresponding to dimensions in the semantic space;

determining the significance of each data point with respect to the predetermined categories; and

constructing a trainable semantic vector for each data point, wherein each trainable semantic vector has dimensions equal to the number of predetermined categories and represents the relative strength of its corresponding data point with respect to each of the predetermined categories; and

combining the trainable semantic vector for each of the at least one data point to form the semantic vector of the sample dataset or the new dataset.

Claims 65-68 (Cancelled)

69. (Currently amended) A computer-readable medium carrying one or more sequences of instructions for classifying new datasets within a predetermined number of categories based on assignment of a plurality of sample datasets to each category, wherein execution of the one or more sequences of instructions by one or more processors causes the one or more processors to perform the machine-executed steps of:

constructing a trainable semantic vector for each sample dataset relative to the predetermined categories in a multi-dimensional semantic space;

constructing a trainable semantic vector for each category based on the trainable semantic vectors for the sample datasets;

receiving a new dataset;

constructing a trainable semantic vector for the new dataset;

determining a distance between the trainable semantic vector for the new dataset and the trainable semantic vector of each category; and

classifying the new dataset within the category whose trainable semantic vector has the shortest distance to the trainable semantic vector of the new ~~dataset.~~dataset;

wherein:

the new data set or each of the sample data sets includes at least one data point;

each data point corresponds to at least one of a word, a phrase, a sentence, a color, a typography, a punctuation, a picture, and a character string; and

the trainable semantic vector for each sample data set or the new dataset is constructed by performing the steps of:

for each data point, identifying a relationship between each data point and predetermined categories corresponding to dimensions in the semantic space;

determining the significance of each data point with respect to the predetermined categories;

constructing a trainable semantic vector for each data point, wherein each trainable semantic vector has dimensions equal to the number of predetermined categories and represents the relative strength of its corresponding data point with respect to each of the predetermined categories; and

combining the trainable semantic vector for each of the at least one data point to form the semantic vector of the sample dataset or the new dataset.

70. (Currently amended) A computer-readable medium carrying one or more sequences of instructions for classifying new datasets within a predetermined number of categories based on assignment of a plurality of sample datasets to each category, wherein execution of the one or more sequences of instructions by one or more processors causes the one or more processors to perform the steps of:

constructing a trainable semantic vector for each sample dataset relative to the predetermined categories in a multi-dimensional semantic space;

receiving a new dataset;

constructing a trainable semantic vector for the new dataset;

identifying a select number of select datasets whose trainable semantic vectors are closest in distance to the trainable semantic vector for the new dataset; and

classifying the new dataset in the category containing the greatest number of the select ~~datasets~~ dataset;

wherein:

the new data set or each of the sample data sets includes at least one data point;

each data point corresponds to at least one of a word, a phrase, a sentence, a color, a typography, a punctuation, a picture, and a character string; and

the trainable semantic vector for each sample data set or the new dataset is constructed by performing the steps of:

for each data point, identifying a relationship between each data point and predetermined categories corresponding to dimensions in the semantic space;

determining the significance of each data point with respect to the predetermined categories;

constructing a trainable semantic vector for each data point, wherein each trainable semantic vector has dimensions equal to the number of predetermined categories and represents the relative strength of its corresponding data point with respect to each of the predetermined categories; and

combining the trainable semantic vector for each of the at least one data point to form the semantic vector of the sample dataset or the new dataset.

Claims 71-78 (Cancelled)